

**Immunological characterization of papain-induced fragments of Clostridium botulinum type A neurotoxin and interaction of the fragments with brain synaptosomes.**

Kozaki S ; Miki A; Kamata Y; Ogasawara J; Sakaguchi G

Department of Veterinary Science, College of Agriculture, University of Osaka Prefecture, Japan.

Infection and immunity (UNITED STATES) Sep 1989 , 57 (9) p2634-9,  
ISSN 0019-9567 Journal Code: 0246127

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Other Citation Owner: NASA

Record type: Completed

Subfile: INDEX MEDICUS

After treatment of Clostridium botulinum type A neurotoxin with papain, three fragments (Mrs, 101,000, 45,000, and 43,000) were purified by hydrophobic and ion-exchange chromatography with a high-performance liquid chromatographic system. Immunoblotting analyses with monoclonal antibodies showed that the 101,000-dalton fragment consisted of the light chain and a part of the heavy chain (H-1 fragment) linked together by a disulfide bond, and the other two fragments were correlated to the remaining portion of the heavy chain (H-2 fragment). The 45,000- and 43,000-dalton fragments effectively competed for binding of the 125I-labeled neurotoxin to synaptosomes, while no inhibition was observed with the 101,000-dalton fragment. The results indicate that the H-2 fragment interacts with the binding site on the neural membrane. The binding of the neurotoxin was impaired by treatment of synaptosomes with neuraminidase. Incorporation of gangliosides into neuraminidase-treated synaptosomes resulted in the restoration of binding. The results suggest that gangliosides are one of the components of the toxin-binding site.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: Botulinum Toxins--immunology--IM; \*Brain--metabolism--ME; \*Clostridium botulinum --immunology--IM; \*Neurotoxins--immunology--IM; \*Papain--pharmacology--PD; \*Synaptosomes--metabolism--ME; Antibodies, Monoclonal; Antigens, Bacterial--immunology--IM; Binding Sites, Antibody; Binding, Competitive; Botulinum Toxins--metabolism--ME; Botulinum Toxins--pharmacology--PD; Immunoblotting; Mice; Mice, Inbred BALB C; Molecular Weight; Neurotoxins--metabolism--ME; Neurotoxins--pharmacology--PD

CAS Registry No.: 0 (Antibodies, Monoclonal); 0 (Antigens, Bacterial); 0 (Binding Sites, Antibody); 0 (Botulinum Toxins); 0 (Neurotoxins)

Enzyme No.: EC 3.4.22.2 (Papain)

Record Date Created: 19890915

Record Date Completed: 19890915

Type A  
update  
5/28/03  
VGB

06610371 90235864 PMID: 2185020

The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene.

Thompson D E ; Brehm J K; Oultram J D; Swinfield T J; Shone C C; Atkinson T; Melling J; Minton N P

Division of Biotechnology, Centre for Applied Microbiology and Research, Porton Down, England.

European journal of biochemistry / FEBS (GERMANY, WEST) Apr 20 1990 , 189 (1) p73-81, ISSN 0014-2956 Journal Code: 0107600

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

A 26-mer oligonucleotide probe was synthesized (based on the determined amino acid sequence of the N-terminus of the *Clostridium botulinum* type A neurotoxin, BoNT/A) and used in Southern blot analysis to construct a restriction map of the region of the clostridial genome encompassing BoNT/A. The detailed information obtained enabled the cloning of the structural gene as three distinct fragments, none of which were capable of directing the expression of a toxic molecule. The central portion was cloned as a 2-kb PvuII-TaqI fragment and the remaining regions of the light chain and heavy chain as a 2.4-kb ScaI-TaqI fragment and a 3.4-kb HpaI-PvuII fragment, respectively. The nucleotide sequence of all three fragments was determined and an open reading frame identified, composed of 1296 codons corresponding to a polypeptide of 149 502 Da. The deduced amino acid sequence exhibited 33% similarity to tetanus toxin, with the most highly conserved regions occurring between the N-termini of the respective heavy chains. Conservation of Cys residues flanking the position at which the toxins are cleaved to yield the heavy chain and light chain allowed the tentative identification of those residues which probably form the disulphide bridges linking the two toxin subfragments.

Tags: Comparative Study

Descriptors: **Botulinum** Toxins--genetics--GE; \**Clostridium botulinum* --genetics--GE; \*Genes, Structural, Bacterial; \*Neurotoxins--genetics--GE; Amino Acid Sequence; Base Sequence; Cloning, Molecular; *Escherichia coli* --genetics--GE; Molecular Sequence Data; Oligonucleotide Probes --biosynthesis--BI; Sequence Homology, Nucleic Acid

Molecular Sequence Databank No.: GENBANK/X52066; GENBANK/UNKNOWN

CAS Registry No.: 0 (Botulinum Toxins); 0 (Neurotoxins); 0 (Oligonucleotide Probes)

Record Date Created: 19900606

Record Date Completed: 19900606

SEQ  
type A

**Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and around the nicking site.**

**DasGupta B R ; Dekleva M L**

Food Research Institute, University of Wisconsin, Madison 53706.

Biochimie (FRANCE) Sep 1990 , 72 (9) p661-4, ISSN 0300-9084

Journal Code: 1264604

Contract/Grant No.: NS17742; NS; NINDS; NS24545; NS; NINDS

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

Clostridium botulinum synthesizes the type A botulinum neurotoxin (NT) as a approximately 150 kDa single **chain** protein. Post-translational proteolytic processing yields a approximately 150 kDa dichain protein composed of a approximately 50 kDa light and approximately 100 kDa heavy **chain** , which has higher toxicity. Trypsin's action mimics the endogenous proteolytic processing. The proteolytic cleavages could occur at 4 sites. We have examined 2 such sites and defined the peptide sequences before and after proteolytic processing. The N-terminal residues of the newly synthesized approximately 150 kDa single **chain** NT, Pro-Phe-Val-Asn-Lys-, remain intact at the N-terminus of the approximately 50 kDa light **chain** generated either in the clostridial culture or in vitro with trypsin or with a protease purified from the homologous bacterial culture. The clostridial protease cleaves the single **chain** NT in vitro, at 1/3 the distance from its N-terminus, on the amino side of Gly of the sequence -Gly-Tyr-Asn-Lys-Ala-Leu-Asn-Asp-Leu- before cleaving the bond Lys-Ala at a slower rate. The data indicate that the dichain NT is formed in the bacterial culture in at least 2 steps. Cleavage at X-Gly produces a approximately 100 kDa heavy **chain** -like fragment which is then truncated; cleavage 4 residues downstream at Lys-Ala, and excision of the tetrapeptide Gly-Tyr-Asn-Lys, generates the mature heavy **chain** with Ala as its N-terminal residue. The approximately 100 kDa heavy **chain** generated in vitro, by nicking the single **chain** NT with trypsin, also has Ala-Leu-Asn- as the N-terminal residues.

Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: \*Amino Acid Sequence; \*Botulinum Toxins--chemistry--CH; \*Clostridium botulinum; Botulinum Toxins--genetics--GE; Clostridium botulinum--genetics--GE; Molecular Sequence Data; Protein Processing, Post-Translational

CAS Registry No.: 0 (Botulinum Toxins)

Record Date Created: 19910311

Record Date Completed: 19910311

**Covalent structure of botulinum neurotoxin type A: location of sulfhydryl groups, and disulfide bridges and identification of C-termini of light and heavy chains.**

Kriegelstein K G ; DasGupta B R; Henschen A H

Department of Molecular Biology and Biochemistry, University of California, Irvine 92717.

**Journal of protein chemistry** (UNITED STATES) Jan 1994 , 13 (1)  
p49-57, ISSN 0277-8033 Journal Code: 8217321

Contract/Grant No.: NS17742; NS; NINDS

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

Botulinum neurotoxin Type A is synthesized by *Clostridium botulinum* as a approximately 150 kD single chain polypeptide. The posttranslational processing of the 1296 amino acid residue long gene product involves removal of the initiating methionine, formation of disulfide bridges, and limited proteolysis (nicking) by the bacterial protease(s). The mature dichain neurotoxin is made of a approximately 50-kD light chain and a approximately 100-kD heavy chain connected by a disulfide bridge. DNA derived amino acid sequence predicted a total of 9 Cys residues (Binz et al., 1990, J. Biol. Chem. 265, 9153-9158; Thompson et al., 1990, Eur. J. Biochem. 189, 73-81). Treatment of the dichain neurotoxin, dissolved in 6 M guanidine. HCl, with 4-vinylpyridine converted 5 Cys residues into S-pyridylethyl cysteine residues; but alkylation after mercaptolysis converted all 9 Cys residues in the S-pyridylethylated form. After confirming the predicted number of Cys residues by amino acid analysis, the positions of the 5 Cys residues carrying sulfhydryl groups and the 4 involved in disulfide bridges were determined by comparing the elution patterns in reversed-phase HPLC of the cyanogen bromide mixtures of the exclusively alkylated and the mercaptolyzed-alkylated neurotoxin. The chromatographically isolated components were identified by N-terminal amino acid sequence analysis. The HPLC patterns showed characteristic differences. The Cys residues predicted in positions 133, 164, 790, 966, and 1059 were found in the sulfhydryl form; Cys 429 and 453 were found disulfide-bridge connecting the light and heavy chains, and Cys 1234 and 1279 were found in an intrachain disulfide-bridge near the C-terminus in the heavy chain. (ABSTRACT TRUNCATED AT 250 WORDS)

Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.


Descriptors: \*Botulinum Toxins--chemistry--CH; Amino Acid Sequence; Botulinum Toxins--biosynthesis--BI; Chromatography, High Pressure Liquid; *Clostridium botulinum*--metabolism--ME; Cyanogen Bromide; Disulfides --analysis--AN; Macromolecular Systems; Molecular Sequence Data; Neurotoxins--chemistry--CH; Peptide Fragments--chemistry--CH; Peptide Fragments--isolation and purification--IP; Sulfhydryl Compounds--analysis --AN; Trypsin

CAS Registry No.: 0 (Botulinum Toxins); 0 (Disulfides); 0 (Macromolecular Systems); 0 (Neurotoxins); 0 (Peptide Fragments); 0 (Sulfhydryl Compounds); 506-68-3 (Cyanogen Bromide)

Enzyme No.: EC 3.4.21.4 (Trypsin)

Record Date Created: 19940726

Record Date Completed: 19940726

 <a href="#">ExPASy Home page</a>	<a href="#">Site Map</a>	<a href="#">Search ExPASy</a>	<a href="#">Contact us</a>	<a href="#">Swiss-Prot</a>
<a href="#">Hosted by NCSC US</a>		<a href="#">Mirror sites:</a> <a href="#">Canada</a> <a href="#">China</a> <a href="#">Korea</a> <a href="#">Switzerland</a> <a href="#">Taiwan</a>		
Search <input type="text" value="Swiss-Prot/TrEMBL"/>		for <input type="text"/>	<input type="button" value="Go"/>	<input type="button" value="Clear"/>

# NiceProt View of Swiss-Prot: P10844

[Printer-friendly view](#)[Quick BlastP search](#)

[\[General\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#)  
[\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

*Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.*

## General information about the entry

Entry name	BXB_CLOBO
Primary accession number	P10844
Secondary accession number	P10843
Entered in Swiss-Prot in	Release 11, July 1989
Sequence was last modified in	Release 26, July 1993
Annotations were last modified in	Release 42, September 2003
<b>Name and origin of the protein</b>	
Protein name	Botulinum neurotoxin type B [Precursor]
Synonyms	EC <u>3.4.24.69</u> BoNT/B Bontoxilysin B
Gene name	BOTB
From	<u>Clostridium botulinum</u> [TaxID: <u>1491</u> ]
Taxonomy	<u>Bacteria</u> ; <u>Firmicutes</u> ; <u>Clostridia</u> ; <u>Clostridiales</u> ; <u>Clostridiaceae</u> ; <u>Clostridium</u> .

## References

- [1] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=92384550; PubMed=1514783; [NCBI, ExPASy, EBI, Israel, Japan]  
Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T., Minton N.P.;  
"Molecular cloning of the Clostridium botulinum structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence."  
Appl. Environ. Microbiol. 58:2345-2354(1992).
- [2] SEQUENCE OF 35-245 FROM NUCLEIC ACID.  
**STRAIN**=NCTC 7273;  
Szabo E.A., Pemberton J.M., Desmarchelier P.M.;  
Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
- [3] SEQUENCE OF 633-993 FROM NUCLEIC ACID.  
**STRAIN**=NCTC 7273;  
MEDLINE=94013372; PubMed=8408542; [NCBI, ExPASy, EBI, Israel,

	<p><u>Japan]</u></p> <p><u>Campbell K., East A.K., Collins M.D.:</u></p> <p>"Gene probes for identification of the botulinal neurotoxin gene and specific identification of neurotoxin types B, E, and F.";</p> <p><u>J. Clin. Microbiol. 31:2255-2262(1993).</u></p>
[4]	<p>SEQUENCE OF <u>1-44</u> AND <u>441-466</u>.</p> <p><b>STRAIN=657;</b></p> <p>MEDLINE=89000987; PubMed=3139097; [<u>NCBI</u>, <u>ExPASy</u>, <u>EBI</u>, <u>Israel</u>, <u>Japan]</u></p> <p><u>Dasgupta B.R., Datta A.:</u></p> <p>"Botulinum neurotoxin type B (strain 657): partial sequence and similarity with tetanus toxin.";</p> <p><u>Biochimie 70:811-817(1988).</u></p>
[5]	<p>SEQUENCE OF <u>1-16</u> AND <u>441-458</u>.</p> <p><b>STRAIN=OKRA;</b></p> <p>MEDLINE=85197963; PubMed=3888113; [<u>NCBI</u>, <u>ExPASy</u>, <u>EBI</u>, <u>Israel</u>, <u>Japan]</u></p> <p><u>Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.:</u></p> <p>"Partial amino acid sequences of botulinum neurotoxins types B and E.";</p> <p><u>Arch. Biochem. Biophys. 238:544-548(1985).</u></p>
[6]	<p>IDENTIFICATION AS ZINC-PROTEASE.</p> <p>MEDLINE=93054694; PubMed=1429690; [<u>NCBI</u>, <u>ExPASy</u>, <u>EBI</u>, <u>Israel</u>, <u>Japan]</u></p> <p><u>Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.:</u></p> <p>"Botulinum neurotoxins are zinc proteins.";</p> <p><u>J. Biol. Chem. 267:23479-23483(1992).</u></p>
[7]	<p>IDENTIFICATION OF SUBSTRATE.</p> <p>MEDLINE=93063293; PubMed=1331807; [<u>NCBI</u>, <u>ExPASy</u>, <u>EBI</u>, <u>Israel</u>, <u>Japan]</u></p> <p><u>Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P., Dasgupta B.R., Montecucco C.:</u></p> <p>"Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.";</p> <p><u>Nature 359:832-835(1992).</u></p>

## Comments

**FUNCTION:** BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF SYNAPTOBREVIN-2.

**CATALYTIC ACTIVITY:** Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.

**COFACTOR:** Binds 1 zinc ion per subunit (*By similarity*).

**SUBUNIT:** DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDiate CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.

**SUBCELLULAR LOCATION:** Secreted.

**MISCELLANEOUS:** THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

**SIMILARITY:** BELONGS TO PEPTIDASE FAMILY M27.

## Copyright

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

## Cross-references



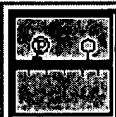
EMBL	M81186; AAA23211.1; [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] -. [ <a href="#">CoDingSequence</a> ] Z11934; CAA77991.1; [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] -. [ <a href="#">CoDingSequence</a> ] X70817; CAA50148.1; [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] -. [ <a href="#">CoDingSequence</a> ]
PIR	A48940; A48940.
PDB	1EPW; 01-NOV-00. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1F31; 01-NOV-00. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1F82; 16-AUG-00. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1F83; 16-AUG-00. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1FQH; 06-DEC-00. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1G9A; 13-NOV-02. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1G9B; 13-NOV-02. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1G9C; 13-NOV-02. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1G9D; 13-NOV-02. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] 1I1E; 21-NOV-01. [ <a href="#">ExpASy</a> / <a href="#">RCSB</a> ] <a href="#">Detailed list of linked structures.</a>
MEROPS	<a href="#">M27.002</a> ; -.
InterPro	<a href="#">IPR000395</a> ; Bontoxilysin. <a href="#">IPR006025</a> ; Zn_MTpeptdse. <a href="#">Graphical view of domain structure.</a>
Pfam	<a href="#">PF01742</a> ; Peptidase_M27; 1.
PRINTS	<a href="#">PR00760</a> ; BONTOXILYSIN.
ProDom	<a href="#">PD001963</a> ; Bontoxilysin; 1. <a href="#">[Domain structure / List of seq. sharing at least 1 domain]</a>
PROSITE	<a href="#">PS00142</a> ; ZINC_PROTEASE; 1.
HOBACGEN	<a href="#">[Family / Alignment / Tree]</a>
BLOCKS	<a href="#">P10844</a> .
ProtoNet	<a href="#">P10844</a> .
ProtoMap	<a href="#">P10844</a> .
PRESAGE	<a href="#">P10844</a> .
DIP	<a href="#">P10844</a> .

ModBase	<a href="#">P10844</a> .
SWISS-2DPAGE	<a href="#">Get region on 2D PAGE</a> .

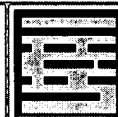
## Keywords

**Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; 3D-structure.**

## Features



[Feature table viewer](#)



[Feature aligner](#)

Key	From	To	Length	Description
INIT_MET	0	0		
CHAIN	<u>1</u>	<u>440</u>	440	BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
CHAIN	<u>441</u>	<u>1290</u>	850	BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
METAL	<u>229</u>	<u>229</u>		ZINC (CATALYTIC) ( <i>BY SIMILARITY</i> ).
ACT_SITE	<u>230</u>	<u>230</u>		<i>BY SIMILARITY</i> .
METAL	<u>233</u>	<u>233</u>		ZINC (CATALYTIC) ( <i>BY SIMILARITY</i> ).
DISULFID	<u>436</u>	<u>445</u>		INTERCHAIN ( <i>PROBABLE</i> ).
CONFLICT	<u>29</u>	<u>29</u>		T -> M (IN REF. <u>4</u> ).
CONFLICT	<u>217</u>	<u>217</u>		R -> G (IN REF. <u>2</u> ).
CONFLICT	<u>224</u>	<u>224</u>		A -> S (IN REF. <u>2</u> ).
CONFLICT	<u>463</u>	<u>463</u>		S -> R (IN REF. <u>4</u> ).

## Sequence information

Length: <b>1290 AA</b> [This is the length of the unprocessed precursor]	Molecular weight: <b>150670 Da</b> [This is the MW of the unprocessed precursor]	CRC64: <b>D21746E2C024DF43</b> [This is a checksum on the sequence]
---	--	--

10	20	30	40	50	60
PVTINNFNYN	DPIDNNNIIM	MEPPFARGTG	RYKAFKITD	RIWIIPERYT	FGYKPEDFNK
70	80	90	100	110	120
SSGIFNRDVC	EYYDPDYLNT	NDKKNIFLQT	MIKLFNRIKS	KPLGEKLLEM	IINGIPYLGD
130	140	150	160	170	180
RRVPLEEFNT	NIASVTVNKL	ISNPGEVERK	KGIFANLIIF	GPGPVLNENE	TIDIGIQNHF

190	200	210	220	230	240
ASREGFGGIM	QMKFCPEYVS	VFNNVQENKG	ASIFNRRGYF	SDPALILMHE	LIHVLHGLYG
250	260	270	280	290	300
IKVDDLPIVP	NEKKFFMQST	DAIQAEELYT	FGGQDPSIIT	PSTDKSIYDK	VLQNFRGIVD
310	320	330	340	350	360
RLNKVLVCIS	DPNININIK	NKFKDKYKFV	EDSEGKYSID	VESFDKLYKS	LMFGFTETNI
370	380	390	400	410	420
AENYKIKTRA	SYFSDSLPPV	KIKNLLDNEI	YTIEEGFNIS	DKDMEKEYRG	QNKAINKQAY
430	440	450	460	470	480
EEISKEHLAV	YKIQMCKSVK	APGICIDVDN	EDLFFIADKN	SFSDDLKNE	RIEYNTQSNY
490	500	510	520	530	540
IENDFPINEL	ILDTDLISKI	ELPSENTESL	TDFNVDVPVY	EKQPAIKKIF	TDENTIFQYL
550	560	570	580	590	600
YSQTFPLDIR	DISLTSSFDD	ALLFSNKVYS	FFSMDYIKTA	NKVVEAGLFA	GWVKQIVNDF
610	620	630	640	650	660
VIEANKSNTM	DKIADISLIV	PYIGLALNVG	NETAKGNFEN	AFEIAGASIL	LEFIPELLIP
670	680	690	700	710	720
VVGAFLESY	IDNKNKIKT	IDNALTKRNE	KWSDMYGLIV	AQWLSTVNTQ	FYTIKEGMYK
730	740	750	760	770	780
ALNYQAQALE	EIIKYRYNIY	SEKEKSNINI	DFNDINSKLN	EGINQAIDNI	NNFINGCSVS
790	800	810	820	830	840
YLMKKMIPLA	VEKLDFDNT	LKKNLLNYID	ENKLYLIGSA	EYEKSKVNKY	LKTIMPFDLIS
850	860	870	880	890	900
IYTNDTILIE	MFNKYNSEIL	NNIILNLRYK	DNNLIDLSGY	GAKVEVYDGV	ELNDKNQFKL
910	920	930	940	950	960
TSSANSKIRV	TQNQNIIFNS	VFLDFSVSFW	IRIPKYKNDG	IQNYIHNEYT	IINCMKNNSG
970	980	990	1000	1010	1020
WKISIRGNRI	IWTLIDINK	TKSVFFEYNI	REDISEYINR	WFFVTITNNL	NNAKIYINGK
1030	1040	1050	1060	1070	1080
LESNTDIKDI	REVIANGEII	FKLDGDIDRT	QFIWMKYFSI	FNTELSQSNI	EERYKIQSYS
1090	1100	1110	1120	1130	1140
EYLKDFWGNP	LMYNKEYYMF	NAGNKNSYIK	LKKDSPVGEI	LTRSKYNQNS	KYINYRDLVI

1150	1160	1170	1180	1190	1200
GEKFIIRRK	NSQSINDDIV	RKEDIYILDF	FNLNQEWVY	TYKYFKKEE	KLFLAPISDS
1210	1220	1230	1240	1250	1260
DEFYNTIQIK	EYDEQPTYSC	QLLFKKDEES	TDEIGLIGIH	RFYESGIVFE	EYKDYFCISK
1270	1280	1290			
WYLKEVKRKP	YNLKLGCNWQ	FIPKDEGWTE			

P10844 in  
FASTA format

[View entry in original Swiss-Prot format](#)

[View entry in raw text format \(no links\)](#)

[Report form for errors/updates in this Swiss-Prot entry](#)

BLAST [BLAST submission on  
ExPASy/SIB  
or at NCBI \(USA\)](#)



Sequence analysis tools: [ProtParam](#),  
[ProtScale](#), [Compute pI/Mw](#),  
[PeptideMass](#), [PeptideCutter](#),  
[Dotlet](#) (Java)



[ScanProsite](#),  
[MotifScan](#)



Search the [SWISS-MODEL  
Repository](#)



[ExPASy Home page](#)

[Site Map](#)

[Search ExPASy](#)

[Contact us](#)

[Swiss-Prot](#)

[Hosted by NCSC US](#)

[Mirror sites:](#)

[Canada](#)

[China](#)

[Korea](#)

[Switzerland](#)

[Taiwan](#)

Search  for

# Swiss-Prot: BXB\_CLOBO

The section of the sequence BXB\_CLOBO (P10844) you have selected corresponds to:

SEQUENCE 1 44

In one-letter code:

1	11	21	31	41	51	
1 PVTINNFNYN	DPIDNNNIIM	MEPPFARGTG	RYKAFKITD	RIWIIPERYT	FGYKPEDFNK	60
61 SSGIFNRDVC	EYYDPDYLNT	NDKKNIFLQT	MIKLFNRIKS	KPLGEKLLEM	IINGIPYLGD	120
121 RRVPLEEFNT	NIASVTVNKL	ISNPGEVERK	KGIFANLIIF	GPGPVLNENE	TIDIGIQNHF	180
181 ASREGFGGIM	QMKFCPEYVS	VFNNVQENKG	ASIFNRRGYF	SDPALILMHE	LIHVLHGLYG	240
241 IKVDDLPIVP	NEKKFFMQST	DAIQAEELYT	FGGQDPSIIT	PSTDKSIYDK	VLQNFRGIVD	300
301 RLNKVLVCIS	DPNININIYK	NKFKDKYKFV	EDSEGKYSID	VESFDKLYKS	LMFGFTETNI	360
361 AENYKIKTRA	SYFSDSLPPV	KIKNLLDNEI	YTIEEGFNIS	DKDMEKEYRG	QNKAINQQAY	420
421 EEISKEHLAV	YKIQMCKSVK	APGICIDVDN	EDLFFIADKN	SFSDDLKNE	RIEYNTQSNY	480
481 IENDFPINEL	ILDIDLISKI	ELPSENTESL	TDFNVDVPVY	EKQPAIKKIF	TDENTIFQYL	540
541 YSQTFFPLDIR	DISLTSSFDD	ALLFSNKVYS	FFSMDYIKTA	NKVVEAGLFA	GWVKQIVNDF	600
601 VIEANKSNTM	DKIADISLIV	PYIGLALNVG	NETAKGNFEN	AFEIAGASIL	LEFIPELLIP	660
661 VVGAFLLLESY	IDNKNKIIKT	IDNALTKRNE	KWSDMYGLIV	AQWLSTVNTQ	FYTIKEGMYK	720
721 ALNYQAQALE	EIIKYRYNIY	SEKEKSNINI	DFNDINSKLN	EGINQAIDNI	NNFINGCSVS	780
781 YLMKKMIPLA	VEKLLDFDNT	LKKNLLNYID	ENKLYLIGSA	EYEKSKVNKY	LKTIMPFDLN	840
841 IYTNDTILIE	MFNKYNSEIL	NNIILNLRYK	DNNLIDLSGY	GAKVEVDYDV	ELNDKNQFKL	900
901 TSSANSKIRV	TQNQNIIFNS	VFLDFSVSFW	IRIPKYKNDG	IQNYIHNEYT	IINCMKNSNG	960
961 WKISIRGNRI	IWTLIDINGK	TKSVFFEYNI	REDISEYINR	WFFVTITNNL	NNAKIYINGK	1020
1021 LESNTDIKDI	REVIANGEII	FKLDGDIDRT	QFIWMKYFSI	FNTELSQSNI	EERYKIQSYS	1080
1081 EYLKDFWGNP	LMYNKEYYMF	NAGNKNSYIK	LKKDSPVGEI	LTRSKYNQNS	KYINYRDLYI	1140
1141 GEKFIIIRKS	NSQSINDDIV	RKEDYIYLDI	FNLNQEWVRY	TYKYFKKEEE	KLFLAPISDS	1200
1201 DEFYNTIQIK	EYDEQPTYSC	QLLFKKDEES	TDEIGLIGIH	RFYESGIVFE	EYKDYFCISK	1260
1261 WYLKEVKRKP	YNLKLGCNWQ	FIPKDEGWTE				

In three-letter code:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1 Pro	Val	Thr	Ile	Asn	Asn	Phe	Asn	Tyr	Asn	Asp	Pro	Ile	Asp	Asn	15
16 Asn	Asn	Ile	Ile	Met	Met	Glu	Pro	Pro	Phe	Ala	Arg	Gly	Thr	Gly	30
31 Arg	Tyr	Tyr	Lys	Ala	Phe	Lys	Ile	Thr	Asp	Arg	Ile	Trp	Ile	Ile	45
46 Pro	Glu	Arg	Tyr	Thr	Phe	Gly	Tyr	Lys	Pro	Glu	Asp	Phe	Asn	Lys	60
61 Ser	Ser	Gly	Ile	Phe	Asn	Arg	Asp	Val	Cys	Glu	Tyr	Tyr	Asp	Pro	75
76 Asp	Tyr	Leu	Asn	Thr	Asn	Asp	Lys	Lys	Asn	Ile	Phe	Leu	Gln	Thr	90
91 Met	Ile	Lys	Leu	Phe	Asn	Arg	Ile	Lys	Ser	Lys	Pro	Leu	Gly	Glu	105

106	Lys	Leu	Leu	Glu	Met	Ile	Ile	Asn	Gly	Ile	Pro	Tyr	Leu	Gly	Asp	120
121	Arg	Arg	Val	Pro	Leu	Glu	Glu	Phe	Asn	Thr	Asn	Ile	Ala	Ser	Val	135
136	Thr	Val	Asn	Lys	Leu	Ile	Ser	Asn	Pro	Gly	Glu	Val	Glu	Arg	Lys	150
151	Lys	Gly	Ile	Phe	Ala	Asn	Leu	Ile	Ile	Phe	Gly	Pro	Gly	Pro	Val	165
166	Leu	Asn	Glu	Asn	Glu	Thr	Ile	Asp	Ile	Gly	Ile	Gln	Asn	His	Phe	180
181	Ala	Ser	Arg	Glu	Gly	Phe	Gly	Gly	Ile	Met	Gln	Met	Lys	Phe	Cys	195
196	Pro	Glu	Tyr	Val	Ser	Val	Phe	Asn	Asn	Val	Gln	Glu	Asn	Lys	Gly	210
211	Ala	Ser	Ile	Phe	Asn	Arg	Arg	Gly	Tyr	Phe	Ser	Asp	Pro	Ala	Leu	225
226	Ile	Leu	Met	His	Glu	Leu	Ile	His	Val	Leu	His	Gly	Leu	Tyr	Gly	240
241	Ile	Lys	Val	Asp	Asp	Leu	Pro	Ile	Val	Pro	Asn	Glu	Lys	Lys	Phe	255
256	Phe	Met	Gln	Ser	Thr	Asp	Ala	Ile	Gln	Ala	Glu	Glu	Leu	Tyr	Thr	270
271	Phe	Gly	Gly	Gln	Asp	Pro	Ser	Ile	Ile	Thr	Pro	Ser	Thr	Asp	Lys	285
286	Ser	Ile	Tyr	Asp	Lys	Val	Leu	Gln	Asn	Phe	Arg	Gly	Ile	Val	Asp	300
301	Arg	Leu	Asn	Lys	Val	Leu	Val	Cys	Ile	Ser	Asp	Pro	Asn	Ile	Asn	315
316	Ile	Asn	Ile	Tyr	Lys	Asn	Lys	Phe	Lys	Asp	Lys	Tyr	Lys	Phe	Val	330
331	Glu	Asp	Ser	Glu	Gly	Lys	Tyr	Ser	Ile	Asp	Val	Glu	Ser	Phe	Asp	345
346	Lys	Leu	Tyr	Lys	Ser	Leu	Met	Phe	Gly	Phe	Thr	Glu	Thr	Asn	Ile	360
361	Ala	Glu	Asn	Tyr	Lys	Ile	Lys	Thr	Arg	Ala	Ser	Tyr	Phe	Ser	Asp	375
376	Ser	Leu	Pro	Pro	Val	Lys	Ile	Lys	Asn	Leu	Leu	Asp	Asn	Glu	Ile	390
391	Tyr	Thr	Ile	Glu	Glu	Gly	Phe	Asn	Ile	Ser	Asp	Lys	Asp	Met	Glu	405
406	Lys	Glu	Tyr	Arg	Gly	Gln	Asn	Lys	Ala	Ile	Asn	Lys	Gln	Ala	Tyr	420
421	Glu	Glu	Ile	Ser	Lys	Glu	His	Leu	Ala	Val	Tyr	Lys	Ile	Gln	Met	435
436	Cys	Lys	Ser	Val	Lys	Ala	Pro	Gly	Ile	Cys	Ile	Asp	Val	Asp	Asn	450
451	Glu	Asp	Leu	Phe	Phe	Ile	Ala	Asp	Lys	Asn	Ser	Phe	Ser	Asp	Asp	465
466	Leu	Ser	Lys	Asn	Glu	Arg	Ile	Glu	Tyr	Asn	Thr	Gln	Ser	Asn	Tyr	480
481	Ile	Glu	Asn	Asp	Phe	Pro	Ile	Asn	Glu	Leu	Ile	Leu	Asp	Thr	Asp	495
496	Leu	Ile	Ser	Lys	Ile	Glu	Leu	Pro	Ser	Glu	Asn	Thr	Glu	Ser	Leu	510
511	Thr	Asp	Phe	Asn	Val	Asp	Val	Pro	Val	Tyr	Glu	Lys	Gln	Pro	Ala	525
526	Ile	Lys	Lys	Ile	Phe	Thr	Asp	Glu	Asn	Thr	Ile	Phe	Gln	Tyr	Leu	540
541	Tyr	Ser	Gln	Thr	Phe	Pro	Leu	Asp	Ile	Arg	Asp	Ile	Ser	Leu	Thr	555
556	Ser	Ser	Phe	Asp	Asp	Ala	Leu	Leu	Phe	Ser	Asn	Lys	Val	Tyr	Ser	570
571	Phe	Phe	Ser	Met	Asp	Tyr	Ile	Lys	Thr	Ala	Asn	Lys	Val	Val	Glu	585
586	Ala	Gly	Leu	Phe	Ala	Gly	Trp	Val	Lys	Gln	Ile	Val	Asn	Asp	Phe	600
601	Val	Ile	Glu	Ala	Asn	Lys	Ser	Asn	Thr	Met	Asp	Lys	Ile	Ala	Asp	615
616	Ile	Ser	Leu	Ile	Val	Pro	Tyr	Ile	Gly	Leu	Ala	Leu	Asn	Val	Gly	630
631	Asn	Glu	Thr	Ala	Lys	Gly	Asn	Phe	Glu	Asn	Ala	Phe	Glu	Ile	Ala	645
646	Gly	Ala	Ser	Ile	Leu	Leu	Glu	Phe	Ile	Pro	Glu	Leu	Leu	Ile	Pro	660
661	Val	Val	Gly	Ala	Phe	Leu	Leu	Glu	Ser	Tyr	Ile	Asp	Asn	Lys	Asn	675
676	Lys	Ile	Ile	Lys	Thr	Ile	Asp	Asn	Ala	Leu	Thr	Lys	Arg	Asn	Glu	690
691	Lys	Trp	Ser	Asp	Met	Tyr	Gly	Leu	Ile	Val	Ala	Gln	Trp	Leu	Ser	705
706	Thr	Val	Asn	Thr	Gln	Phe	Tyr	Thr	Ile	Lys	Glu	Gly	Met	Tyr	Lys	720
721	Ala	Leu	Asn	Tyr	Gln	Ala	Gln	Ala	Leu	Glu	Glu	Ile	Ile	Lys	Tyr	735
736	Arg	Tyr	Asn	Ile	Tyr	Ser	Glu	Lys	Glu	Lys	Ser	Asn	Ile	Asn	Ile	750
751	Asp	Phe	Asn	Asp	Ile	Asn	Ser	Lys	Leu	Asn	Glu	Gly	Ile	Asn	Gln	765
766	Ala	Ile	Asp	Asn	Ile	Asn	Asn	Phe	Ile	Asn	Gly	Cys	Ser	Val	Ser	780
781	Tyr	Leu	Met	Lys	Lys	Met	Ile	Pro	Leu	Ala	Val	Glu	Lys	Leu	Leu	795
796	Asp	Phe	Asp	Asn	Thr	Leu	Lys	Lys	Asn	Leu	Leu	Asn	Tyr	Ile	Asp	810
811	Glu	Asn	Lys	Leu	Tyr	Leu	Ile	Gly	Ser	Ala	Glu	Tyr	Glu	Lys	Ser	825
826	Lys	Val	Asn	Lys	Tyr	Leu	Lys	Thr	Ile	Met	Pro	Phe	Asp	Leu	Ser	840
841	Ile	Tyr	Thr	Asn	Asp	Thr	Ile	Leu	Ile	Glu	Met	Phe	Asn	Lys	Tyr	855
856	Asn	Ser	Glu	Ile	Leu	Asn	Asn	Ile	Ile	Leu	Asn	Leu	Arg	Tyr	Lys	870
871	Asp	Asn	Asn	Leu	Ile	Asp	Leu	Ser	Gly	Tyr	Gly	Ala	Lys	Val	Glu	885
886	Val	Tyr	Asp	Gly	Val	Glu	Leu	Asn	Asp	Lys	Asn	Gln	Phe	Lys	Leu	900
901	Thr	Ser	Ser	Ala	Asn	Ser	Lys	Ile	Arg	Val	Thr	Gln	Asn	Gln	Asn	915
916	Ile	Ile	Phe	Asn	Ser	Val	Phe	Leu	Asp	Phe	Ser	Val	Ser	Phe	Trp	930
931	Ile	Arg	Ile	Pro	Lys	Tyr	Lys	Asn	Asp	Gly	Ile	Gln	Asn	Tyr	Ile	945
946	His	Asn	Glu	Tyr	Thr	Ile	Ile	Asn	Cys	Met	Lys	Asn	Asn	Ser	Gly	960
961	Trp	Lys	Ile	Ser	Ile	Arg	Gly	Asn	Arg	Ile	Ile	Trp	Thr	Leu	Ile	975
976	Asp	Ile	Asn	Gly	Lys	Thr	Lys	Ser	Val	Phe	Phe	Glu	Tyr	Asn	Ile	990
991	Arg	Glu	Asp	Ile	Ser	Glu	Tyr	Ile	Asn	Arg	Trp	Phe	Phe	Val	Thr	1005
1006	Ile	Thr	Asn	Asn	Leu	Asn	Asn	Ala	Lys	Ile	Tyr	Ile	Asn	Gly	Lys	1020
1021	Leu	Glu	Ser	Asn	Thr	Asp	Ile	Lys	Asp	Ile	Arg	Glu	Val	Ile	Ala	1035
1036	Asn	Gly	Glu	Ile	Ile	Phe	Lys	Leu	Asp	Gly	Asp	Ile	Asp	Arg	Thr	1050
1051	Gln	Phe	Ile	Trp	Met	Lys	Tyr	Phe	Ser	Ile	Phe	Asn	Thr	Glu	Leu	1065
1066	Ser	Gln	Ser	Asn	Ile	Glu	Glu	Arg	Tyr	Lys	Ile	Gln	Ser	Tyr	Ser	1080

```

1081 Glu Tyr Leu Lys Asp Phe Trp Gly Asn Pro Leu Met Tyr Asn Lys 1095
1096 Glu Tyr Tyr Met Phe Asn Ala Gly Asn Lys Asn Ser Tyr Ile Lys 1110
1111 Leu Lys Lys Asp Ser Pro Val Gly Glu Ile Leu Thr Arg Ser Lys 1125
1126 Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr Arg Asp Leu Tyr Ile 1140
1141 Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn Ser Gln Ser Ile 1155
1156 Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr Leu Asp Phe 1170
1171 Phe Asn Leu Asn Gln Glu Trp Arg Val Tyr Thr Tyr Lys Tyr Phe 1185
1186 Lys Lys Glu Glu Glu Lys Leu Phe Leu Ala Pro Ile Ser Asp Ser 1200
1201 Asp Glu Phe Tyr Asn Thr Ile Gln Ile Lys Glu Tyr Asp Glu Gln 1215
1216 Pro Thr Tyr Ser Cys Gln Leu Leu Phe Lys Lys Asp Glu Glu Ser 1230
1231 Thr Asp Glu Ile Gly Leu Ile Gly Ile His Arg Phe Tyr Glu Ser 1245
1246 Gly Ile Val Phe Glu Glu Tyr Lys Asp Tyr Phe Cys Ile Ser Lys 1260
1261 Trp Tyr Leu Lys Glu Val Lys Arg Lys Pro Tyr Asn Leu Lys Leu 1275
1276 Gly Cys Asn Trp Gln Phe Ile Pro Lys Asp Glu Gly Trp Thr Glu

```

---

Direct similarity search submission of this subsequence to

BLAST

BLAST submission on  
ExPASy/SIB  
 or at NCBI (USA)



Sequence analysis tools:

ProtParam, ProtScale, Compute  
pI/Mw, PeptideMass,  
PeptideCutter, Dotlet (Java)



ScanProsite



Direct Submission to  
SWISS-MODEL



ExPASy Home page

Site Map

Search ExPASy

Contact us

Swiss-Pr t

updated  
5/03  
[Signature]

ID CBBONT standard; DNA; PRO; 4041 BP.  
 XX  
 AC M81186;  
 XX  
 SV M81186.1  
 XX  
 DT 28-MAY-1992 (Rel. 32, Created)  
 DT 04-MAR-2000 (Rel. 63, Last updated, Version 4)  
 XX  
 DE Clostridium botulinum neurotoxin type B (botB) gene, complete cds.  
 XX  
 KW botB gene; neurotoxin type B.  
 XX  
 OS Clostridium botulinum  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 XX  
 RN [1]  
 RP 1-4041  
 RA Whelan S.M., Elmore M.J, Bodsworth N.J., Brehm J.K., Atkinson T.,  
 RA Minton N.P.;  
 RT "Complete nucleotide sequence of the Clostridium botulinum gene encoding  
 RT the type B neurotoxin";  
 RL Unpublished.  
 XX  
 DR GOA; P10844; P10844.  
 DR SWISS-PROT; P10844; BXB\_CLOBO.  
 XX  
 FH Key Location/Qualifiers  
 FH  
 FT source 1..4041  
 FT /db\_xref="taxon:1491"  
 FT /organism="Clostridium botulinum"  
 FT CDS 57..3932  
 FT /codon\_start=1  
 FT /db\_xref="GOA:P10844"  
 FT /db\_xref="SWISS-PROT:P10844"  
 FT /transl\_table=11  
 FT /function="vertebrate neurotoxin"  
 FT /product="neurotoxin type B"  
 FT /gene="botB"  
 FT /protein\_id="AAA23211.1"  
 FT /translation="MPVTINNFNNDPIDNNNIIMMEPPFARGTGRIYKAFKITDRIWI  
 FT IPERYTFGYKPEDFNKSSGIFNRDVCEYDPPDYLNTNDKKNIFLQTMIKLFNRKSKPL  
 FT GEKLEMIINGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGP  
 FT GPVLNENETIDIGIQNHFASREGFGGIMQMKFCPEYVSFNNVQENKGASIFNRRGYFS  
 FT DPALILMHLELIHVLHGLYGIKVDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIIT  
 FT PSTDKSIYDKVLQNFGRGIVDRNLKVLVCISDPNINININIKNFKDKYKFVEDSEGKYSI  
 FT DVESFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIIEGFN  
 FT ISDKDMEKEYRGONKAINKQAYEEISKEHLAVYKIQMCKSVKAPGICIDVDNEDLFFIA  
 FT DKNSFSDDLKNERIEYNTQSNYIENDFPINELILDTDLISKIELPSENTESLTDPNVD  
 FT VPVYEKQPAIKKIFTDENTIFQYLYSQTFFPLDIRDISLTSSFDALLFSNKVYSFFSMD  
 FT YIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETA  
 FT KGNFENAFEIAGASILLEFIPELLIPVVGAFLLSYIDNKNKIIKTIDNALTKRNEKWS  
 FT DMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDF  
 FT NDINSKLNENGINQAIDNINNFINCGSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDE  
 FT NKLYLIGSAEYKSKVNKYLKTIMPFDLISYTNDTILIEFMKNYSEILNIIILNLRYK  
 FT DNNLIDLSGYGAKVEVDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSF  
 FT WIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLDINGKTKSVFFEY  
 FT NIREDISSEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDI  
 FT DRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKN  
 FT SYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRKSNSSQSIINDDIVRKEDY  
 FT IYLDFFNLNQEWRYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLF  
 FT KKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIP  
 FT KDEGWTE"



XX  
SQ

Sequence 4041 BP; 1679 A; 383 C; 645 G; 1334 T; 0 other;

tgcgcat	ttta	tgggc	atttaa	aaggg	atata	aacttaaa	at	atatttatgc	60
cagttaca	aat	aat	at	at	at	at	at	attattatga	120
tggagc	ctcc	atttgc	gaga	ggtac	gggga	gatattata	aa	atcacagatc	180
gtattt	ggat	aatacc	ggaa	agata	tactt	ttggatata	aa	tttaataaaa	240
gttccg	gtat	tttta	ataga	gatgtt	tgtg	aatattatga	tc	ttaaatacta	300
atgata	aaaa	gaata	tattt	ttaca	aacaa	tgatcaagt	tt	atcaaatcaa	360
aaccatt	ggg	tga	aaagt	ttaga	gatga	ttataaat	gg	cttggagata	420
gacgtg	ttcc	actcga	agag	ttta	acacaa	acattgctag	tg	taaaatttaa	480
tcagta	atcc	aggaga	agtg	gagcg	aaaaa	aaggtatttt	cg	caaattta	540
gacctg	ggcc	agttt	taaat	gaaa	atgaga	ctatagatat	ag	gtatacaa	600
catca	aggg	aggct	tcggg	ggtata	atgc	aaatgaagt	tg	cccagaa	660
tattt	ataa	tg	tcaaga	aacaa	aggcg	caagtata	tt	aatagacgt	720
cagatcc	cagc	cttgat	atta	atgcat	gaac	ttatacatgt	tt	tacatgga	780
ttaa	agtaga	tgattt	tacca	attgt	tacc	atgaaaaaa	at	ttttttatg	840
atgctata	caga	agaa	ctata	tacat	ttgg	aggaca	ag	atcccagc	900
cttctac	gga	taaa	agtatc	tatga	taag	ttttgcaaaa	tt	tttagagg	960
gactta	acaa	gg	ttttag	tgc	at	atccta	acat	taata	1020
ataa	attt	taa	agata	aat	tcgt	g	g	g	1080
tagaa	agttt	tgata	aaatta	tata	aaag	ct	ta	atgtttg	1140
cagaaa	atta	taaa	ataaaa	actag	agctt	cttatttt	tag	tttcc	1200
aaata	aaaaa	ttt	attagat	aatga	aatct	atactata	ga	ggaagg	1260
ataa	agatat	ggaaa	agaa	tata	gaggtc	agaata	aa	gc	1320
aagaa	attag	caagg	agcat	ttgg	ctgtat	ataagata	ca	aatgtg	1380
ctcc	aggaat	atgtat	tgat	gttga	taatg	aagattt	gtt	ctttatag	1440
gttttt	caga	tgattt	tatct	aaaa	cga	gaataga	ata	taata	1500
tagaaa	atga	cttcc	ctata	aatga	atttaa	ttttagata	c	tgatt	1560
aattac	caag	tga	aaata	caca	gaatc	actt	ctg	attt	1620
aaaa	acaacc	cgctata	aaaa	aaa	attttt	ta	cagat	gaaaa	1680
actctc	agac	atttcc	tcta	gatata	agag	atata	agttt	aacat	1740
cattatt	tatt	ttcta	acaaa	gtttat	tcat	tttttt	ctat	ggatt	1800
ataa	agtgg	aga	gcagga	ttatt	tgcag	gttgg	gtga	acagata	1860
taatc	gaagc	taata	aaagc	aata	ctatgg	ataaa	attgc	agata	1920
cttat	atag	attag	cttta	aatgt	aggaa	atgaa	acagc	taaa	1980
ctttt	tgagat	tg	caggagcc	agtatt	ctac	taga	atttat	accaga	2040
tagtt	ggagc	ctttt	tatta	gaat	catata	ttgaca	ataa	aaata	2100
tagata	atgc	ttta	actaaa	agaa	atgaaa	aatgg	agtga	tatgt	2160
cgca	atggct	ctca	acagtt	aata	ctcaat	tttata	caat	aaa	2220
cttta	aat	tca	gcacaa	gcatt	ggaag	aaata	ataa	ata	2280
ctgaaa	aaga	aa	agtcaaat	atta	acatcg	attt	ta	atga	2340
aggg	tattaa	cca	agctata	gata	aatata	ata	atttt	at	2400
attt	aatgaa	aaaa	atgatt	ccatt	agctg	tagaaa	aat	actag	2460
tca	aaaaaaa	ttt	gttaaat	tata	tagatg	aaa	ataaat	atatt	2520
aatat	gaaaa	atca	aaaagta	aata	aaata	act	tgaaa	accat	2580
tata	tacc	caa	tgata	caata	ctaa	tagaaa	tg	ttta	2640
ata	atattat	ctt	aaattt	agata	ataag	ataa	taatt	at	2700
ggg	caaagg	ag	aggtatat	gat	ggagtcg	agct	ta	atga	2760
ctag	ttcagc	aaat	agtaag	attag	agtga	ctca	aatca	gaat	2820
tg	ttccttga	tttt	agcgtt	agct	tttgg	taaga	atacc	taa	2880
tacaaa	att	tatt	cataat	gaat	atacaa	taatt	aattg	tatg	2940
ggaaa	at	t	atgggg	aat	aggata	tatg	gactt	aat	3000
ccaa	atcggt	at	ttttt	tgaa	tata	acataa	gaga	agatat	3060
gg	ttttttgt	aact	attact	aata	at	tttga	ata	acgctaa	3120
taga	atcaaa	tac	agatat	aa	agata	taa	gaga	agttat	3180
ttaa	attaga	tg	gtgatata	gat	aga	acac	aat	ttttt	3240
ttaa	acgga	at	taag	tcaa	atattg	agaa	agata	taaa	3300
aat	attt	taaa	agattt	ttt	gg	aaat	cctt	ta	3360
atg	cgggg	gaa	taaaa	attca	tata	tt	aaac	taa	3420
taac	acgtag	caa	atata	aat	ctta	aat	atata	taa	3480
gagaaa	aat	tatt	ataaga	agaa	gtcaa	att	ctca	atc	3540
gaaa	agaaga	ttat	atat	at	at	tt	aat	ttta	3600
cctata	aaata	ttt	taagaaa	gag	gaagaaa	aat	gt	ttt	3660
atgag	ttt	ta	actata	caa	ataaaa	aat	at	gatga	3720
agtt	gctttt	taaaa	aa	gat	ga	ctgat	gagat	aggat	3780

```
gtttctacga atctggaatt gtatttgaag agtataaaga ttatttttgt ataagtaaatt 3840
ggtacttaaa agaggtaaaa aggaaaccat ataatttaaa attgggatgt aattggcagt 3900
ttattcctaa agatgaaggg tggactgaat aatataacta tatgctcagc aaacctattt 3960
tatataagaa aagtttaagt ttataaaatc ttaagtttaa ggatgtagct aaattttgaa 4020
tattagataa actacatggt t 4041
```

//